

Received April 23, 2020, accepted May 7, 2020, date of publication May 18, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2995302

# Spectrum-Weighted Tensor Discriminant Analysis for Motor Imagery-Based BCI

SHOULIN HUANG<sup>ID</sup><sup>1</sup>, YANG CHEN<sup>1</sup>, TONG WANG<sup>ID</sup><sup>1</sup>, (Member, IEEE),  
AND TING MA<sup>1,2,3,4</sup>, (Member, IEEE)

<sup>1</sup>School of Electronic and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, Guangdong 518055, China

<sup>2</sup>Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing 100069, China

<sup>3</sup>National Clinical Research Center for Geriatric Disorders, Xuanwu Hospital, Capital Medical University, Beijing 510510, China

<sup>4</sup>Artificial Intelligence Research Center, Peng Cheng Laboratory, Shenzhen 518055, China

Corresponding author: Ting Ma (tma@hit.edu.cn)

This study was supported by grants from National Key Research and Development Program of China (2018YFC1312000), the National Natural Science Foundation of China (61801145), and the Basic Research Project of Shenzhen Science and Technology Program (JCYJ20180306171800589).

**ABSTRACT** Spatio-temporal filtering has been widely used for extracting discriminative features in the motor imagery-based brain-computer interface (MI-BCI). In order to obtain high performance, the algorithms need to enhance robustness or find class-discriminative bands for the spatial filter. However, the existing methods either cannot derive the spatial and spectral filters with a unique objective function for guaranteeing convergence or rarely consider the combined optimization of spatial-spectral filters and other patterns for enhancing the discrimination. In this study, we present a novel feature extraction method termed Spectrum-weighted Tensor Discriminant Analysis (SwTDA), which optimizes spectral filters along with spatial filters and other associated patterns by tensor-based discriminant analysis. The proposed method considers intrinsic spatial-spectral-temporal information contained by the physiological signal and hence can identify discriminative characteristics robustly. The effectiveness of the algorithm is demonstrated by comparing it with several state-of-the-art methods on two datasets involving 15 different subjects. Results indicate that the SwTDA method yields higher classification accuracies than the competing methods. Furthermore, interpretable spatial-spectral patterns that are determined by the algorithm can be used for further analysis of the MI-based EEG signal.

**INDEX TERMS** Brain computer interface (BCI), spatio-spectral filter, tensor-based discriminant analysis, motor imagery.

## I. INTRODUCTION

Brain-computer interface (BCI) system can translate the user's intent to computer commands, so it is valuable for people with severe motor disabilities [1], [2]. Electroencephalogram (EEG) is often used in many BCI studies for its noninvasive measurement and low cost. One type of EEG-based BCI system is based on the modulation of sensorimotor rhythm (SMR), which follows the event-related (de)synchronization (ERD/ERS) when the subjects imagine movements of their limbs, i.e., motor imagery (MI) [3]. Three simple limb motor imagery (sMI) tasks (i.e., left hand, right hand, and foot) are typical for the application because of the discriminative characteristics between each other. In order to obtain more output commands in MI-based BCI for the need of many real-world applications, it is necessary to develop

The associate editor coordinating the review of this manuscript and approving it for publication was Juan Wang<sup>ID</sup>.

new types of MI tasks. Yi *et al.* [5] have investigated the differences of the induced brain oscillatory patterns between sMI and compound limb motor imagery (cMI).

In recent years, many new classification techniques have been designed and explored to classify the EEG signal. Adaptive classifiers (e.g., [37]) are demonstrated to be generally superior to static ones, even with unsupervised adaptation; Riemannian geometry-based methods (e.g., [6], [36]) have reached state-of-the-art performances on multiple BCI problems; deep learning methods (e.g., [32]–[34]) have been applied to enhance the EEG classification accuracy. Other advanced techniques, such as transfer learning [11] or boosting ensemble methods [9], [35], are also believed to be able to improve the performance. However, MI-based BCIs, the same as most pattern recognition, not only use a classifier but also apply feature extraction method to obtain characteristics as discriminative as possible.

Spatio-temporal filters have been proven to be successful in enhancing the signal-to-noise ratio of the highly noisy and low-spatial-resolution EEG signals [7], [14], [19], [23], [25], [30]. The well-known common spatial patterns (CSP) [8] is a highly successful spatial filtering technique to extract Motor imagery-based signal features. However, good performance of CSP severely depends on performing temporal filtering with a proper frequency band, which is important for portraying SMR activities. Researches (e.g., [10], [30]) have shown that responsive frequency bands of ERD/ERS are not consistent with inter- and intra-subjects. Many recent works have been proposed and aim to resolve the problem. The strategies they used are to enhance robustness or find class-discriminative bands for spatial filtering. These works can be categorized into three schemes as follows (to avoid confusion, we use the temporal or spectral filter on the aspect of class-discriminative bands, whereas the temporal pattern for extracting time domain-related features).

The first scheme focuses on boosting robust to the individual variability by using regularization approaches, robust data averaging, and/or new divergence measures [29]–[31]. Regularization of the covariance matrix, e.g., [29], is one common approach to increase robustness, especially in small-sample settings. Recently, beta divergence-based CSP algorithms (e.g., [30] and [31]) incorporate different types of regularization schemes into the optimization process and improve the performance robustly. The extended common spatial pattern (ECSP) [4], which aims to obtain enough variance information from the prior knowledge available data (e.g., the mean data matrices of the two classes), has been reported to achieve high performance in the dataset IIb of BCI competition IV. However, these methods still neglect the frequency information, and hence cannot provide to portray the rhythmic activities in the sense of signal analysis.

The second category passes the signal through a filter bank and then finds spatial filters from the filtered EEG signal. The Filter Bank Common Spatial Pattern (FBCSP) [14] and its variants (e.g., SFBCSP [15] and SCSSP [16]) belong to this category. Although these methods have been reported to achieve high performance, the spectral range of each subband must be predetermined. As an improvement, Bayesian spatio-spectral filter optimization (BSSFO) [17] is proposed to devise the arbitrary frequency bands which are generated by a particle-based approximation method. More recently, an error-correcting output coding-based CSP (ECCSP-TB2B) is proposed by Shahtalebi and Mohammadi [19]. In the part of feature extraction, the ECCSP-TB2B extends the BSSFO technique by using two frequency bands within each particle. It is worth pointing out for the BSSFO and ECCSP-TB2B that the spatial weights are optimized by CSP and the frequency bands are found by an optimization problem based on the approximate estimation of the posterior probability density function. This means that the cost functions for these two optimization problems are different.

The third category (e.g., [23]–[26]) performs the simultaneous optimization of spatial and spectral filters. The spatio-temporally regularized CSP (STRCSP) [23] embeds several time-delayed signals into raw EEG signals, and obtains spatial and temporal bandpass filters simultaneously based on CSP computation. However, the small time-delayed number limits the flexibility of the spectral filters. The discriminative filter bank CSP (DFBCSP) [25] conducts the optimization of bandpass temporal filters and spatial filters by solving two subproblems of generalized eigenvalue decomposition sequentially and alternatively. Along with a different line, the iterative spatio-spectral patterns learning (ISSPL) [24] optimizes temporal filters equivalently in the frequency domain by alternate iteration. At each iteration, the ISSPL computes spatial filters by the CSP technique, whereas the spectral filters are equivalently solved by the max-margin machine of a classifier. In other words, two different criteria are respectively used for spatial and spectral optimization. As an improvement on ISSPL, the maximizing mutual information of spatial spectral features (MMISS) [26] utilizes the criterion of maximizing the mutual information instead of the strategy ISSPL uses. However, as mentioned by the authors, MMISS is limited in finding the global optimum because the convergence is related to a gradient search method.

Overall, in order to obtain the class-discriminative bands, it is preferred to optimize spectral filters rather than using a fixed filter bank. As mentioned above, most of the studies derive the spatial and spectral filters alternately by two different optimization criteria. The alternating iteration is hard to guarantee to be converged because the cost functions for these two problems are different. Furthermore, these algorithms rarely consider the optimization to combine the spatial-spectral filters and other patterns. Motivated by overcoming the major issues of the methods mentioned above, we develop a new algorithm, named Spectrum-weighted Tensor Discriminant Analysis (SwTDA), for EEG classification. EEG data are firstly transformed into the time-frequency domain for forming tensor-based samples by adding the frequency mode. Then the SwTDA finds the spectral filters with the spatial filters and other associated patterns (e.g., temporal pattern) based on Fisher's discrimination criterion from the repeatedly calculated residues of the original tensor data. Our method belongs to the third category and contributes to three primary aspects.

First, because of processing EEG signals in the form of multiway arrays, the proposed method can find spectral filters together with spatial and temporal patterns. Some recent researches (e.g., [34]) have shown that temporal information can contribute to high classification accuracy. The decomposition that takes into account spatial-spectral-temporal information and their consistency for trials can provide the advantage of capturing interactions and couplings among various variables (e.g., temporal, spectral, and spatial components) often with physical or physiological meanings and

interpretations [7]. Hence features extracted by the SwTDA preserve intrinsic interactive information between multi-dimensions, which can help for boosting the robustness.

Second, the SwTDA yields the frequency mode through time-frequency transform so that it can find spectral weights directly in the frequency domain. This contrasts with the STRCSP, where the low order filters limit the flexibility in shaping the passband and stopband. It also differs from the BSSFO and ECCSP-TB2B, where the parameters of spectral filters are defined as cutoff frequencies and thus the passband and other frequency characteristics are only decided by the used filter.

Third, the SwTDA achieves the simultaneous optimization of spatial-spectral filters and other patterns based on maximizing tensor-based Fisher's discrimination criterion, which has a single cost function and has been proved to guarantee monotonic convergence in reference [20]. Other methods, such as ISSPL and ECCSP-TB2B, do not preserve this characteristic because the objective functions for optimizing the spatial and spectral filters differ from each other.

The remainder of the paper is organized as follows: the mathematical formulation of the proposed SwTDA method is presented in Section II. The experimental results come on Section III. Then, the discussions are carried out in Section IV. Finally, Section V concludes the paper.

## II. SPECTRUM-WEIGHTED TENSOR DISCRIMINANT ANALYSIS (SwTDA)

### A. NOTATIONS AND BASIC TENSOR CONCEPTS

We denote tensors (i.e., multi-way arrays) by boldface Euler script letters. For example,  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is a  $N$ -order (or mode) tensor, where  $I_n (n = 1, \dots, N)$  is the size of the mode- $n$ . Matrices are denoted by boldface capital letters, e.g.,  $X \in R^{I \times J}$ ; vectors are denoted by boldface lowercase letters, e.g.,  $x$ ; scalars are denoted by lowercase letters, e.g.,  $x$ . The mode- $n$  product of a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  with a matrix  $A \in R^{J \times I_n}$ , is defined as

$$\mathcal{Y} = \mathcal{X} \times_n A \quad (1)$$

where  $\mathcal{Y}_{i_1, \dots, i_{n-1}, j, i_{n+1}, \dots, i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1, i_2, \dots, i_N} A_{j, i_n}$  and  $A_{j, i_n}$  denotes the element  $(j, i_n)$  of the matrix  $A$ .

The product of a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  with a set of matrices  $\{A^{(n)} \in R^{J_n \times I_n}, n = 1, \dots, N\}$  is denoted as

$$\mathcal{Y} = \mathcal{X} \times \{A\} = \mathcal{X} \times_1 A^{(1)} \dots \times_N A^{(N)} \quad (2)$$

Elementwise, we have

$$\begin{aligned} \mathcal{Y}_{j_1, \dots, j_N} \\ = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1, i_2, \dots, i_N} A_{j_1, i_1}^{(1)} \dots A_{j_N, i_N}^{(N)} \end{aligned} \quad (3)$$

### B. FEATURE EXTRACTION

Assume that we have a training set,  $\{\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2}, m = 1, 2, \dots, M\}$ , where  $I_1$  is the number of channels and  $I_2$  is

**TABLE 1.** List of symbols in this paper.

Name	Description
$M$	total sample number of training set
$\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}$	tensor of $m$ th-sample, $m \in 1, 2, \dots, M$
$\mathcal{X}_{m,f}$	the $f$ th spatio-temporal matrix, i.e., $\mathcal{X}_{m,f} = \mathcal{X}_m(:, :, f)$
$\mathcal{X}_m(:, n, f)$	the $n$ th column of $\mathcal{X}_{m,f}$
$\mathcal{X}_m(c, n, f)$	element $(c, n)$ of $\mathcal{X}_{m,f}$
$p \in \mathbb{R}^F$	spectral weight vector
$U \in \mathbb{R}^{I_1 \times R}$	spatial projection matrix
$W \in \mathbb{R}^{I_2 \times J}$	temporal projection matrix
$w_j$	the $j$ th column of $W$
$u_r$	the $r$ th column of the $U$
$Z_m \in \mathbb{R}^{R \times J}$	projection of $m$ th sample $\mathcal{X}_m$ through $\{S, W, p\}$
$z_{r,j}$	projection of $m$ th sample $\mathcal{X}_m$ through $\{s_r, w_j, p\}$

the number of sample points, and  $M$  is the total trial number. In order to optimize spectral filter directly in the frequency domain, each trial data is represented as a 3-order tensor with modes (*channel*  $\times$  *time*  $\times$  *frequency*) by wavelet transformation. To be more specific, the tensor  $\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  of the  $m$ -th trial is given by the amplitude of wavelet transformed signal

$$\mathcal{X}_m = \|\psi * X_m\| \quad (4)$$

where  $\psi$  is the wavelet function. We select the complex Morlet wavelet,  $\psi(t) = \frac{1}{\sqrt{\pi f_b}} \exp(i2\pi f_c t) \exp(-\frac{t^2}{f_b})$ , as the mother wavelet. The bandwidth parameter  $f_b$  is set to 2, and the wavelet center frequency  $f_c$  is set to 1, since these parameters have been successfully applied in the time-frequency analysis of EEG signal [22]. If the broad frequency range we focus on is  $[f_1, f_2]$ , the mode size of the frequency domain,  $I_3$ , is decided by

$$I_3 = \text{size}([f_1 : f_r : f_2]) \quad (5)$$

where  $f_r$  is frequency resolution;  $[f_1 : f_r : f_2]$  denotes the vector  $[f_1, f_1 + f_r, f_1 + 2f_r, \dots, f_2]$ ;  $\text{size}(\cdot)$  return the element number of the vector. Note that  $f_r = 1$  Hz is used in this study.

The tensor  $\mathcal{X}_m$  consists of  $I_3$  spatio-temporal matrices, each of which denotes a frequency component. Given a spectral filter  $p = [p_1, p_2, \dots, p_{I_3}]^T \in \mathbb{R}^{I_3}$ , we define the spectral projection as the weighted sum of the matrices:

$$Y_m = \sum_{f=1}^{I_3} p_f \mathcal{X}_{m,f} \quad (6)$$

where  $\mathcal{X}_{m,f} = \mathcal{X}_m(:, :, f), f \in \{1, 2, \dots, I_3\}$ .  $p_f$  denotes the weight that  $\mathcal{X}_{m,f}$  contributes to feature.

A spatial projection (named spatial filters) and a temporal projection are used to reduce the dimension of the spectral projection  $Y_m$ . Given a spatial filter  $u_r \in \mathbb{R}^{I_1}$  and a temporal

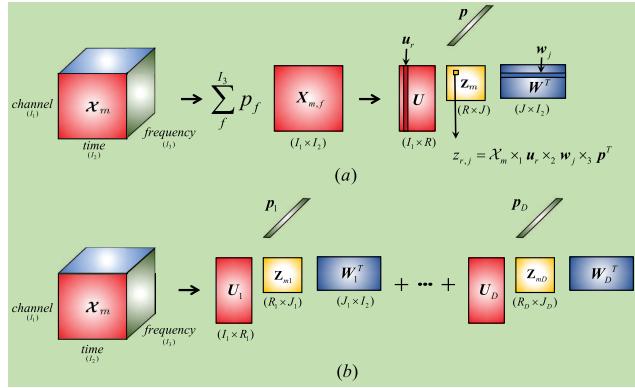
pattern  $\mathbf{w}_j \in \mathbb{R}^{I_2}$ , the spatial-temporal projection can be formulated as

$$z_{r,j} = \mathbf{u}_r^\top \mathbf{Y}_m \mathbf{w}_j \quad (7)$$

where  $z_{r,j}$  is the projection signal of  $\mathcal{X}_m$  by  $\{\mathbf{u}_r, \mathbf{w}_j, \mathbf{p}\}$ . Finally, the combination of the spatial, temporal, and spectral projections can be written as the product of the tensor  $\mathcal{X}_m$  with  $\{\mathbf{U}, \mathbf{W}, \mathbf{p}\}$

$$\mathbf{Z}_m = \mathcal{X}_m \times_1 \mathbf{U}^\top \times_2 \mathbf{W}^\top \times_3 \mathbf{p}^\top \quad (8)$$

where  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_R]$ ,  $R$  is the number of spatial filters;  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_J]$ ,  $J$  is the number of temporal projection;  $\mathbf{Z}_m$  consists of the projection features and  $z_{r,j}$  is the element  $(r, j)$  of  $\mathbf{Z}_m$ . Fig. 1(a) illustrates the principle of SwTDA.



**FIGURE 1.** Graphical explanation of the SwTDA projection model.  
(a) SwTDA projects  $\mathcal{X}_m$  to  $\mathbf{Z}_m$  through  $\{\mathbf{U}, \mathbf{W}, \mathbf{p}\}$ . (b) Projection into terms with  $D$  spectral filters.

Fisher's discrimination criterion (FDC) can be adopted to find  $\{\mathbf{U}, \mathbf{W}, \mathbf{p}\}$ . The reason is that maximizing FDC has a single cost function for the optimization problems and guarantees monotonic convergence [20]–[22]. The scatter matrices and the corresponding FDC can be defined as:

$$S_w = \sum_{m=1}^M \|\mathbf{Z}_m - \bar{\mathbf{Z}}_{k_m}\|^2, S_b = \sum_{k=1}^K M_k \|\bar{\mathbf{Z}}_k - \bar{\mathbf{Z}}\|^2 \quad (9)$$

$$J = \arg \max_{\{\mathbf{U}, \mathbf{W}, \mathbf{p}\}} S_b / S_w \quad (10)$$

where  $S_w$  and  $S_b$  are within-class and between-class scatter matrices respectively;  $K$  is the number of classes;  $k_m$  is the class label for the  $m$ -th training sample;  $\bar{\mathbf{Z}}_{k_m}$  is the mean feature of class  $k_m$ ; and  $\bar{\mathbf{Z}}$  is the mean of all sample feature, i.e.,  $\bar{\mathbf{Z}} = \frac{1}{M} \sum_{m=1}^M \mathbf{Z}_m$ .  $\|\cdot\|^2$  denote  $L_2$  Frobenius norm. There exist no closed-form solutions for (10) to determine  $\{\mathbf{U}, \mathbf{W}, \mathbf{p}\}$  simultaneously, so the alternating projection method that fix all but one is used as a suboptimal solution. Please see Appendix A for details to solve the problem (10).

Assuming that  $\mathbf{p}_1, \dots, \mathbf{p}_D$  are the spectral filters we finally solve for, equation (8) can be rewrote as follows

$$\mathbf{Z}_{m,d} = \mathcal{X}_{m,d} \times_1 \mathbf{U}_d^\top \times_2 \mathbf{W}_d^\top \times_3 \mathbf{p}_d^\top, \quad d = 1, 2, \dots, D \quad (11)$$

### Algorithm 1 Spectrum-Weighted Tensor Discriminant Analysis (SwTDA)

```

1: input:  $M$  training tensor  $\{\mathcal{X}_m \in \mathbb{R}^{I_1 \times I_2 \times I_3}, m = 1, \dots, M\}$ , the number of spectral filters  $D$ ;
2: output: The set of projection matrices  $\{\mathbf{U}_d \in \mathbb{R}^{I_1 \times R_d}, \mathbf{W}_d \in \mathbb{R}^{I_2 \times J_d}, \mathbf{p}_d\}, d = 1, \dots, D$ ;
3: for  $1 \leq d \leq D$  do
4:   initial  $\mathbf{U}_d, \mathbf{W}_d, \mathbf{p}_d$ ;
5:   repeat
6:     fix  $\mathbf{U}_d$  and  $\mathbf{W}_d$  to solve the problem:  $\mathbf{p}_d = \arg \max (\mathbf{S}_b / \mathbf{S}_w)$ ;
7:     fix  $\mathbf{W}_d$  and  $\mathbf{p}_d$  to solve the problem:  $\mathbf{U}_d = \arg \max (\mathbf{S}_b / \mathbf{S}_w)$ ;
8:     fix  $\mathbf{U}_d$  and  $\mathbf{p}_d$  to solve the problem:  $\mathbf{W}_d = \arg \max (\mathbf{S}_b / \mathbf{S}_w)$ ;
9:   until Stop criterion is met
10:  for  $1 \leq m \leq M$  do
11:     $\mathbf{Z}_{m,d} = \mathcal{X}_m \times_1 \mathbf{U}_d^\top \times_2 \mathbf{W}_d^\top \times_3 \mathbf{p}_d^\top$ ;
12:     $\mathcal{X}_m = \mathcal{X}_m - \mathbf{Z}_{m,d} \times_1 \mathbf{U}_d \times_2 \mathbf{W}_d \times_3 \mathbf{p}_d$ ;
13:  end for
14: end for

```

here, we set  $\mathcal{X}_m$  as  $\mathcal{X}_{m,1}$ . And for  $d > 1$ , let  $\mathcal{X}_{m,d} = \mathcal{X}_{m,d-1} - \mathbf{Z}_{m,d-1} \times_1 \mathbf{U}_{d-1} \times_2 \mathbf{W}_{d-1} \times_3 \mathbf{p}_{d-1}$ . This means that we obtain a set of projections with the scatter ratio criterion from the repeatedly calculated residues of  $\mathcal{X}_m$ , and each  $\mathbf{p}_d$  explains the most discriminative frequency pattern of the remaining variation. As shown in Fig. 1(b), the decomposition model of SwTDA can be represented as

$$\mathcal{X}_m = \sum_{d=1}^D \mathbf{Z}_{m,d} \times_1 \mathbf{U}_d \times_2 \mathbf{W}_d \times_3 \mathbf{p}_d \quad (12)$$

The pseudocode of the proposed SwTDA is summarized in Algorithm 1. Note that the SwTDA is easily extended into four or more higher order case. To be more specific, given  $\mathcal{X}_m$  is a  $L$ -order tensor, equation (12) can be rewrote as

$$\mathcal{X}_m = \sum_{d=1}^D \mathcal{Z}_{m,d} \times_1 \mathbf{U}_d^{(1)} \dots \times_{L-1} \mathbf{U}_d^{(L-1)} \times_L \mathbf{p}_d \quad (13)$$

where  $\mathbf{U}_d^{(l)}$  ( $l = 1, \dots, L-1$ ) is the projection matrix in mode- $l$  and  $\mathcal{Z}_{m,d} \in \mathbb{R}^{I_{d,1} \times \dots \times I_{d,L-1}}$  is the feature tensor.

Indeed, if  $D$  is set to 1, SwTDA is a special case of GTDA [20] or HODA [22] with the objective of maximizing the scatter ratio. Furthermore, SwTDA is a heuristic approach with residue calculation, where the spectral filters are obtained one after another. Therefore, there is no orthogonality between any two spectral filters. Additionally, mathematical arguments in reference [20] show that maximizing the scatter differential (or ratio) by the alternating method guarantees monotonic convergence. The MATLAB code of SwTDA is available upon requests from the authors.

### C. FEATURE SELECTION

Features of a trial can be reshaped to form a vector  $\mathbf{v}_m \in R^T$ , where  $T = \sum_{d=1}^D I_{d,1} \times I_{d,2} \dots \times I_{d,L-1}$ . In order to obtain the most discriminative features for classification as well as reduce the feature dimension, Fisher Score [27] is employed for ranking the features. To be specific, given feature matrix  $\mathbf{V} \in R^{M \times T}$  for the  $M$  training samples, let  $\boldsymbol{\mu}$  denote the mean feature vector of all classes whereas  $\boldsymbol{\mu}_k$  and  $\boldsymbol{\sigma}_k$  respectively denote the mean and standard deviation vectors of the  $k$ -class feature vectors. The  $i$ -th Fisher score, which corresponds to the  $i$ -th column of  $\mathbf{V}$ , is computed below:

$$F(i) = \frac{\sum_{k=1}^K M_k (\mu_{k,i} - \mu_i)^2}{\sum_{k=1}^K M_k \sigma_{k,i}^2} \quad (14)$$

where  $\mu_{k,i}$  and  $\mu_i$  are the  $i$ -th element of  $\boldsymbol{\mu}_k$  and  $\boldsymbol{\mu}$  respectively, and  $\sigma_{k,i}$  is the  $i$ -th element of  $\boldsymbol{\sigma}_k$ . Significant features corresponding to the largest  $N$  ( $N \ll T$ ) scores could be chosen for training the classifier. In the testing phase, a trial tensor  $\mathcal{X}_t$  is first projected to the feature components, and then the components are selected and fed into the classifier for predicting the class label.

## III. EXPERIMENTS AND RESULTS

In this section, we compare the proposed method with the following algorithms: (i) the conventional CSP; (ii) the FBCSP method [14]; (iii) the STRCSP method [23], which optimizes bandpass temporal filters with spatial filters in the CSP-based framework; (iv) the divCSP [30], which is based on divergence measures and incorporates robust property into optimization process; (v) the CSP\AM-BA-SVM method [28], which is based on the idea of FBCSP but applies a hybrid Attractor Metagene (AM) and Bat Algorithm (BA) for feature selection along with SVM optimization; (vi) the ECSP method [4], which takes into account prior knowledge available from data and leads to a set of features that enhance the discrimination between the two classes; (vii) the method in reference [34], which employs the FBCSP to select the temporal representation for a convolutional neural network, named as Channel-wise Convolution with Channel Mixing (C2CM); and (viii) the ECCSP-TB2B [19], which iteratively optimizes a set of particles (each particle consisting of 2 bandpass filters) and obtains CSP-based features from the bandpass filtered EEG signal. The experiments are performed based on two public databases. The performance is measured by the classification accuracy, i.e.,  $acc = \frac{M_c}{M_a} \times 100\%$ , where  $M_c$  represents the number of correct classification trials, and  $M_a$  is the total trial number of the testing set.

### A. DATASET ASSESSMENT

Dataset A was from international BCI dataset IVa of the BCI competition III (available at <http://www.bbci.de/competition/iii/>) [13]. EEG signals from five healthy subjects ( $aa, al, av, aw, ay$ ) were recorded at 1000 Hz sampling rate from 118-scalp positions with high-pass and low-pass filters 0.05 and 200 Hz. This dataset consists of two motor

**TABLE 2.** Number of training (labelled) trials “#tr” and test (unlabelled) trials “#te” for each subject.

	<i>aa</i>	<i>al</i>	<i>av</i>	<i>aw</i>	<i>ay</i>
#TR	168	224	84	56	28
#TE	112	56	196	224	252

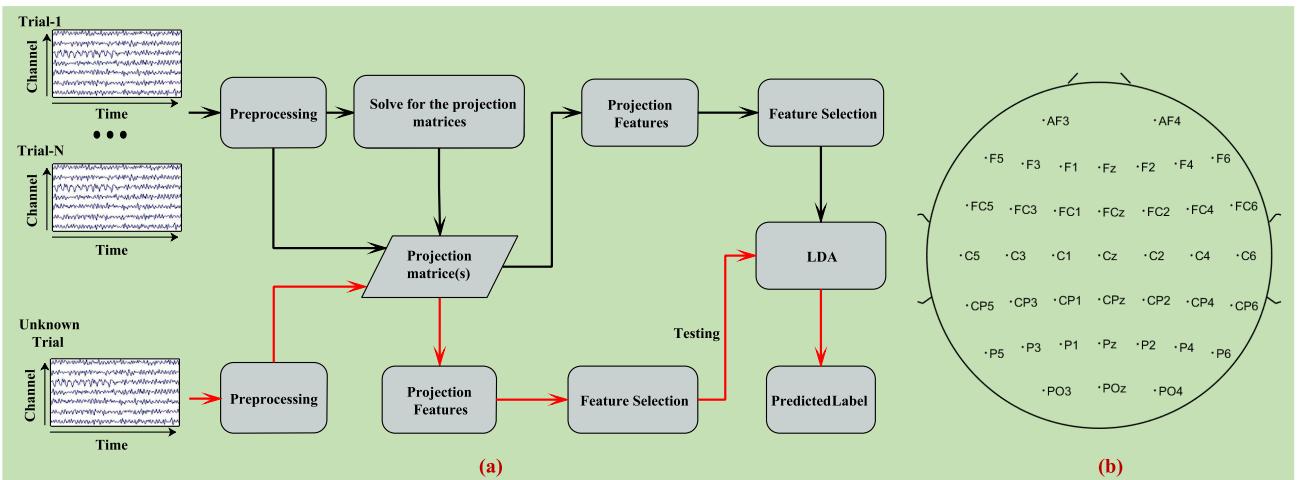
imageries: right hand and right foot. Table 2 shows the respective number of training trials and test trials for each subject. The EEG data were down-sampled to 100 Hz (by picking each 10th sample) in our experiment.

Dataset B was provided by Yi *et al.* [5] (available at <https://doi.org/10.7910/DVN/27306>). This dataset consists of six kinds of motor imagination tasks, i.e., left-hand (LH), right-hand (RH), both-hand (BH), foot (F), left-hand with right-foot (LH&RF), and reft-hand with left-foot (RH&LF). Signals of 64 EEG channels were recorded from ten healthy subjects (S1-S10) with sampling rate of 1000 Hz, and an additional 50 Hz notch filter was used during data acquisition. Similarly, all data were down-sampled to 100 Hz for analysis. A total of 480 trials were collected for each subject (80 trials per task, among which top 60% numbers of trials were used for the training data set, and the remaining trials were for testing). In order to evaluate the single-trial binary classification performance between sMI and cMI tasks, nine binary classification subsets were constructed from the data of each subject: LH vs. BH, LH vs. LH&RF, LH vs. RH&LF, RH vs. BH, RH vs. LH&RF, RH vs. RH&LF, F vs. BH, F vs. LH&RF and F vs. RH&LF. Besides, another three subsets were constructed to evaluate the classification performance between two cMI tasks, i.e., BH vs. LH&RF, BH vs. RH&LF, and LH&RF vs. RH&LF. Thereby, there were 12 binary subsets for each subject, and finally, a total of 120 subsets were obtained.

### B. PARAMETER SELECTION AND EXPERIMENTAL PROCEDURE

Fig. 2(a) shows the general pipeline for all comparative algorithms. Although a wide range of classifiers is available, a simple classifier is preferred so that the performance is mainly contributed by the feature extraction algorithm rather than the classifier. We typically used linear discriminant analysis (LDA) to model and classify MI tasks except for the CSP\AM-BA-SVM and C2CM methods. If there are no special instructions, details of implementation for both datasets are listed as follows. A homogeneous setting of channels and time window is adopted for all the comparative methods for a fair comparison.

In the preprocessing stage, the EEG data of 40 channels (see Fig. 2(b)) were filtered in 7-30 Hz with a fifth-order Butterworth filter (the FBCSP and C2CM were the exceptions for this procedure), and the signals from the period of 0.5-3.5s were selected for each trial. So a trial was represented as a matrix of 40 channels  $\times$  300 time points, and for the SwTDA, the tensor of each trial was size of 40 channels  $\times$  300 time points  $\times$  24 frequency bins according to the equations (4) and (5). Since the number of time points was much



**FIGURE 2.** (a) The general framework for single-trial EEG classification in this study. The black and red arrows stand for the training and testing procedure respectively. (b) 40 electrodes (according to the 10-20 international system) were used in this study.

**TABLE 3.** The accuracy[%] comparison for different methods applied on the test data of Dataset A.

Subjects	CSP	FBCSP	STRCSP	divCSP	CSP\AM-BA [28]	ECSP	C2CM	ECCSP-TB2B	SwTDA
aa	81.25	83.92	60.71	83.04	<b>86.61</b>	80.36	85.71	84.82	<b>86.61</b>
al	98.21	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	98.21	<b>100.0</b>	<b>100.0</b>	98.21
av	60.71	64.80	56.12	<b>70.40</b>	66.84	60.20	68.37	66.33	<b>70.40</b>
aw	87.95	<b>96.42</b>	58.48	86.61	90.63	85.27	94.64	95.09	90.63
ay	74.60	61.90	59.92	83.33	80.95	<b>91.67</b>	75.40	80.56	87.69
Average	80.54	81.40	67.04	84.68	85.01	83.14	84.82	85.36	<b>86.71</b>

larger than the number of channels or frequency bins, the scatter matrices consequently would cause an ill-conditioned eigenvalue problem. Hence, the strategy of multiple time intervals was used in practice. To be specific, each trial was finally reshaped as a 4-D tensor with modes  $40 \text{ channels} \times 15 \text{ time segments} \times 20 \text{ time points} \times 24 \text{ frequency bins}$ . For FBCSP or C2CM, the filter bank was designed to cover the frequency band of 7-31 Hz, which consisted of six nonoverlapping temporal bandpass filters that covered a bandwidth of 4 Hz.

In the stage of feature extraction, three pairs of spatial filters corresponding to the largest and smallest eigenvalues were obtained by CSP, divCSP, STRCSP, ECSP, each sub-band of FBCSP, C2CM, or ECCSP-TB2B. For SwTDA, the number of spectral filters  $D$  was set to 3. For the  $d$ -th spectral filter, we assumed  $J_d = S_d = R_d = 3$ , where  $S_d$  was the number of time segment patterns. Moreover, we also assumed  $J_1 = J_2 = J_3$ .

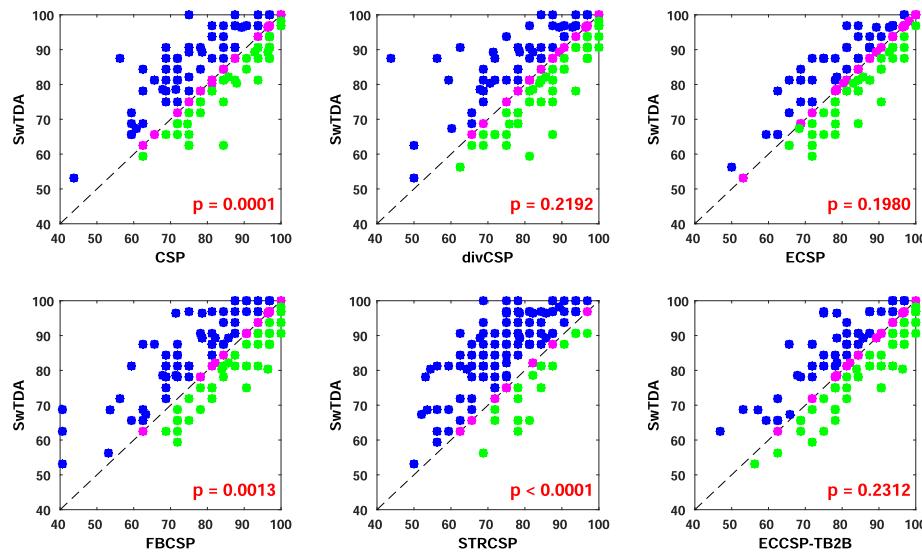
Furthermore, other parameters that we tuned to obtain the best classification accuracy were decided by  $5 \times 5$  cross-validation on the training set for each method and subject (or subset). The dimension of log-variance features was chosen out of {2, 4, 6} for CSP, FBCSP, divCSP, and STRCSP, whereas the number of features  $N$  was selected from 1 to 6 with step 1 in SwTDA. The ECSP method uses linear regression method LASSO to decide the feature dimension. The regularization parameter  $\lambda_1$  in the LASSO

cost function was selected from a set of 200 logarithmically spaced points between  $1e-5$  and 1. In the C2CM, temporal features were extracted from  $2 \times N_s$  spatially filtered channels. The parameter  $N_s$  was chosen out of {1, 2, 3}. The filter order and regularization coefficient in STRCSP were selected from the sets {6, 8, 10, 12, 14, 16} and  $10^{-6} \times \{1, 2, \dots, 50\}$  respectively. In divCSP, the parameter  $\lambda$  was selected from the set of 11 candidates {0, 0.1, 0.2, ..., 1} and beta parameter was chosen out of  $\{-0.0005, -0.001, -0.0015, \dots, 0, 0.5, 1\}$ . For the ECCSP-TB2B method, the number of particles was selected within the range of 5 to 50 with steps of 5.

Besides, the divCSP, ECCSP-TB2B, and SwTDA need a stop criterion, which guarantees a stable convergence. The stopping criterion for iterations of divCSP was that the difference between two successive iterations was less than  $1 \times 10^{-8}$ , or the number of iterations has achieved 1000. The maximum iterative number for the ECCSP-TB2B was varied from 5 to 50 with steps of 5, which was determined by cross-validation. In the case of learning one spectral filter for SwTDA, the iteration stopped when the change of Fisher ratio between two consecutive iterations was less than a preset threshold  $6 \times 10^{-4}$  or iteration number reached 40.

### C. RESULTS FOR DATASET A

Table 3 shows the testing classification accuracies for dataset A. We can see that our algorithm achieves the highest mean accuracy of 86.71%. Although the FBCSP obtains the best



**FIGURE 3.** Scatter plots showing classification accuracies of the SwTDA (y-axis) against six other methods(x-axis). The points above the diagonals indicate the superiority of the algorithm on the y-axis over the one on the x-axis. The p-value of the Wilcoxon signed rank test is shown at the right-bottom corner. The null hypothesis of the Wilcoxon test is that the median of the accuracy rate differences is greater or equal to zero. For  $p < 0.05$ , we reject this null hypothesis; thus, we say that our method significantly outperform the baseline.

performance for subjects *al*, *aw*, it only gives slightly better mean accuracy than CSP, because of weak accuracy for subjects *ay*. It is worth noting that the STRCSP method obtains weak classification performances in this experiment. However, it yields 83.52% of average accuracy when we only modify the filtering band from 7-30 to 4-47 Hz. Moreover, the combination of STRCSP and SFLDA (sparse fisher linear discriminant analysis) classifier has been proved to be effective in [23]. Hence, two likely explanations for the poor performances by STRCSP+LDA include: (1) the STRCSP algorithm constructs for each trial an augmented data that contains both the original EEG signal and  $N - 1$  time-delayed copies, where  $N$  is the order number of the frequency filters. In order to alleviate serious dimensionality issue, fig3 the flexibility of the frequency filters with a small order ( $N \leq 16$ ) is still limited for spectral optimization; and (2) good performance of the STRCSP needs the classifier with sparsity to address the curse-of-dimensionality issue.

As an improvement on FBCSP, the CSP\AM-BA [28], which selects the frequency bands based on the Bat algorithm (BA), is reported to achieve an average accuracy of 85.01%. And the ECCSP-TB2B finds the particles (2 frequency bands per particle) based on a Bayesian framework and yields 85.36% of mean accuracy. We have also compared the proposed method with the C2CM [34], which is considered as one representative method of deep learning applied to EEG-based BCI. The advantage of deep learning is that a large number of hidden layers can be used for extracting features out of input data. In our applications, the C2CM achieves good performances except for the data of subject *ay*, which only has a small number of training samples. As mentioned in [7], [34], one challenge of deep learning is how a small number of samples can be used to train

high capacity networks without suffering from the overfitting issue. Furthermore, an excellent deep learning method still needs to include a data preparation stage, in which a feature extraction method is employed to represent EEG signals in a compact relevant manner without any significant loss in information [32]–[34].

#### D. RESULTS FOR DATASET B

Compound limb motor imagery (cMI) is a more complex cognitive process than simple limb motor imagery (sMI) [18]. Fig. 3 shows the classification accuracies of the testing sets of the 120 subsets. The accuracy rate of our approach is represented on the y-axis, i.e., if the coordinate point is above the diagonal line, then our method outperforms the method represented on the x-axis. The red number at the lower right corner denotes the p-value when applying the one-sided Wilcoxon sign-rank test. The null hypothesis of the Wilcoxon test is that the median of the accuracy rate differences [our method (y-axis)-baseline method (x-axis)] is greater or equal to zero. For  $p < 0.05$ , we reject this null hypothesis; thus, we say that our method significantly outperforms the others.

We can see from the plots that the SwTDA significantly increases classification accuracy rates compared to the CSP ( $p = 0.0001$ ), FBCSP ( $p = 0.0013$ ) and STRCSP ( $p < 0.0001$ ). Although the improvements are not significant, we also can see that the SwTDA performs better than the divCSP ( $p = 0.2192$ ), ECSP ( $p = 0.1980$ ) and ECCSP-TB2B ( $p = 0.2312$ ). The results of 120 subsets are classified into 12 groups according to the task of motor imagery. The mean classification accuracy for each group is provided in Table 4. The results show that the SwTDA achieves the best mean accuracy of 84.4%, which is slightly better than the accuracy of ECCSP-TB2B.

**TABLE 4.** Average accuracies (%)  $\pm$  standard deviation for each group with different methods, tested on Dataset B.

Groups \ Methods	CSP	FBCSP	divCSP	STRCSP	ECSP	ECCSP-TB2B	SwTDA
LH vs BH	81.2 $\pm$ 10.3	75.0 $\pm$ 16.6	75.5 $\pm$ 16.4	70.6 $\pm$ 10.8	80.6 $\pm$ 12.5	76.3 $\pm$ 15.2	<b>82.9</b> $\pm$ 14.7
LH vs LH&RF	76.0 $\pm$ 8.0	78.3 $\pm$ 9.1	76.9 $\pm$ 12.4	72.5 $\pm$ 6.1	76.4 $\pm$ 12.7	79.8 $\pm$ 10.4	<b>79.9</b> $\pm$ 11.5
LH vs RH&LF	88.7 $\pm$ 9.9	<b>91.2</b> $\pm$ 9.7	88.4 $\pm$ 10.9	74.1 $\pm$ 4.9	85.6 $\pm$ 9.1	90.8 $\pm$ 8.2	<b>91.2</b> $\pm$ 8.1
RH vs BH	77.7 $\pm$ 16.7	79.0 $\pm$ 16.9	78.7 $\pm$ 17.0	73.2 $\pm$ 12.8	80.0 $\pm$ 15.5	<b>82.2</b> $\pm$ 12.3	80.9 $\pm$ 15.0
RH vs LH&RF	89.3 $\pm$ 10.8	88.1 $\pm$ 13.6	<b>92.1</b> $\pm$ 7.9	80.4 $\pm$ 8.8	87.5 $\pm$ 7.7	88.4 $\pm$ 13.9	87.8 $\pm$ 10.9
RH vs RH&LF	79.1 $\pm$ 11.2	75.7 $\pm$ 9.1	82.1 $\pm$ 7.3	73.6 $\pm$ 7.0	<b>85.6</b> $\pm$ 6.7	78.8 $\pm$ 6.7	79.9 $\pm$ 7.5
F vs BH	76.2 $\pm$ 11.4	83.0 $\pm$ 10.9	82.0 $\pm$ 10.2	72.6 $\pm$ 11.3	79.7 $\pm$ 10.8	<b>84.0</b> $\pm$ 9.9	81.6 $\pm$ 9.1
F vs LH&RF	84.4 $\pm$ 13.3	85.4 $\pm$ 13.3	85.1 $\pm$ 14.6	83.8 $\pm$ 11.8	86.7 $\pm$ 11.6	87.9 $\pm$ 11.1	<b>90.4</b> $\pm$ 9.3
F vs RH&LF	79.6 $\pm$ 8.0	81.0 $\pm$ 11.8	84.9 $\pm$ 6.8	82.6 $\pm$ 7.8	<b>88.7</b> $\pm$ 7.3	83.8 $\pm$ 12.1	87.5 $\pm$ 11.0
BH vs LH&RF	82.3 $\pm$ 8.4	85.8 $\pm$ 8.5	85.8 $\pm$ 8.5	79.4 $\pm$ 13.6	84.0 $\pm$ 9.9	<b>86.3</b> $\pm$ 8.2	85.4 $\pm$ 8.2
BH vs RH&LF	82.8 $\pm$ 15.6	80.7 $\pm$ 10.2	84.8 $\pm$ 10.4	79.3 $\pm$ 9.3	84.7 $\pm$ 7.6	83.6 $\pm$ 9.6	<b>85.8</b> $\pm$ 8.9
LH&RF vs RH&LF	77.4 $\pm$ 20.8	75.4 $\pm$ 19.0	<b>83.3</b> $\pm$ 13.7	71.9 $\pm$ 10.6	77.7 $\pm$ 15.8	82.4 $\pm$ 16.0	80.2 $\pm$ 15.1
Average	81.2 $\pm$ 12.8	81.5 $\pm$ 13.2	83.3 $\pm$ 12.1	76.2 $\pm$ 10.4	83.1 $\pm$ 11.2	83.7 $\pm$ 11.7	<b>84.4</b> $\pm$ 11.3

## IV. DISCUSSION

### A. PERFORMANCE ANALYSIS

Since the frequency band of ERD/ERS is usually subject-specific, the CSP works well in some subjects but fails at others when the raw EEG signals are bandpass filtered with a fixed frequency range 7-30 Hz. As a further extension of CSP, the FBCSP obtains the EEG rhythms from a filter bank and hence performs slightly better than the CSP. The divCSP encompasses different types of robust properties in the computing process of the spatial filters and hence performs excellently in a broad frequency range. But this also means that it is hard to obtain exact frequency information for portraying rhythmic activities. Different from the CSP, FBCSP, and divCSP, the STRCSP allows finding temporal bandpass filters within the CSP computing. However, as a feature extraction method, the STRCSP suffers from deterioration of performance when combined with a simple LDA classifier in our applications.

The ECSP uses prior knowledge available from data to produce discriminative features and hence significantly outperforms the conventional CSP method. However, similar to divCSP, the ECSP neglects to find the class-discriminative bands, which are crucial for capturing the ERD/ERS change. From the results of two datasets, the SwTDA method that optimizes the spectral filter appears to achieve better performance than the ECSP. In our applications, the ECCSP-TB2B can achieve excellent results in most of the cases. However, deterioration of performance happens on some subsets. During the optimization procedure, the particles (i.e., frequency bands) are updated by adding a disturbance that follows a normal distribution to the values of the band limits. Then the filter bank is checked for continuing the next update or stopping the iteration. Because it is difficult to traverse all the candidates of the cutoff frequencies, the value of the variance in the normal distribution and the number of iterations need to be selected appropriately, or the final bands may be suboptimal in the sense of classification enhancement.

Besides, defining the bandpass filters by a simple band limitation-based way falls into the case that the passband and other frequency characteristics will only be determined by the selected filter. This problem can be demonstrated in the next subsection.

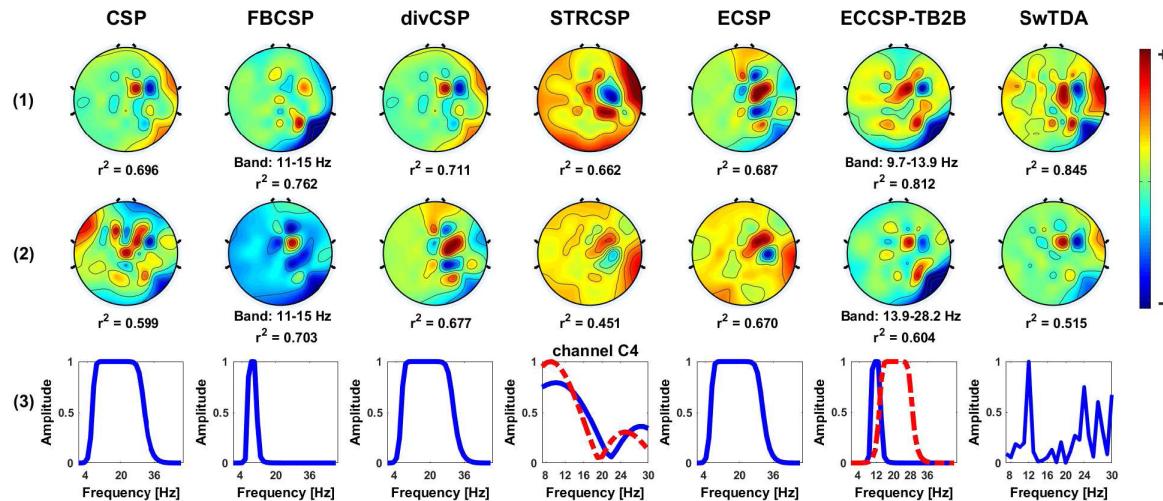
The SwTDA is proposed to find the patterns by tensor-based discriminant analysis, which is a natural extension of LDA. The maximization of Fisher's discrimination finds the spectral filters and the associated spatial-temporal patterns simultaneously. On the one hand, unlike the conventional CSP and FBCSP, the proposed method directly designs the user-specific parameters for the filters such as the passbands. On the other hand, different from ECSP and ECCSP-TB2B, the proposed method has the opportunity to capture the interactions and couplings within multi-dimensions and seems promising to lead to more robust EEG classification. The results confirm that the SwTDA is effective in extracting discriminative features of the ERD/ERS activity.

### B. SPATIAL-SPECTRAL OPTIMIZATION

One advantage of spatial-spectral optimization is that it allows to reveal discriminating parts in the spectrum and topography. To demonstrate differences between the spatial-spectral filters obtained by the comparative methods, one specific subset is picked up for subject S8. During each trial of this subset, the subject performs motor imageries of F and LH&RF. Fig. 4 shows the spatial and spectral filters for this subset. The  $r^2$ -value [13], [21] which is listed beneath each weight map is used to evaluate the discriminative information derived from corresponding spatial pattern. We define the  $r^2$ -value as

$$r^2 = \left( \frac{\sqrt{N_1 * N_2}}{N_1 + N_2} \frac{\text{mean}(z_1) - \text{mean}(z_2)}{\text{std}(z_1 \cup z_2)} \right)^2 \quad (15)$$

where  $z_1$  and  $z_2$  are the features generated from the same spatial filter for class 1 and 2 respectively, and  $N_1$  and  $N_2$  are the sample numbers of calss 1 and class 2 respectively. Larger  $r^2$ -value indicates a higher separability of features.



**FIGURE 4.** The most discriminative patterns obtained by different methods for subset S8-8, which is associated with the MI tasks of F versus LH&RF. The first and second rows are the scalp weight maps corresponds to two of the most discriminative spatial filters. The bottom row (3) displays the spectral characteristics associated with the scalp weight for STRCSP and SwTDA. Specifically for STRCSP and ECCSP-TB2B, the blue solid line and the red dash-dotted line denote the frequency response curves associated with their scalp weight maps of the first and second rows respectively.

We can observe that all the methods have extracted the important spatial components for this specific example. Although there exist similar weight maps between some methods, different  $r^2$ -values are obtained because of adopting different optimization criteria. Features generated from the SwTDA may be more correlated with the imagery tasks for the highest  $r^2$ -value. To be specific, the topographic maps provided by SwTDA show that, channels (i.e., FC2 and FC4) over centro-parietal on the right brain hemisphere show the most influential weights with opposite directions for discrimination respectively, which would explain underlying cortical activity patterns of subject S8 during the motor imagery tasks.

Fig. 4 also provides the spectral information associated with the spatial filters. Note that the CSP, FBCSP, divCSP, ECSP and ECCSP-TB2B obtain frequency response curves by 5-th order Butterworth bandpass filter, and the STRCSP achieves them by the optimized temporal filters whereas the spectral weights of SwTDA denote the amplitude-frequency characteristics directly. We can observe that all the methods have included the important components of higher alpha (**10-16 Hz**) rhythms. But for the scheme that adopts the fixed filter bank strategy (such as the FBCSP), since each frequency band is treated independently, possible correlations between different EEG rhythms may be ignored when it is used for further analysis of the signal. On the other hand, we cannot obtain the spectral characteristics exactly from the curves obtained by the STRCSP, since the bandpass filter with low order limit the flexibility in shaping the passband and stopband. By contrast, the spectral filter obtained by SwTDA clearly exhibits that the important rhythms concentrate on 12 Hz and 24 Hz. This is in accordance with the spectral characteristics presented by the ECCSP-TB2B. However, unlike the ECCSP-TB2B where the passband characteristics are decided by the used filter (e.g., Butterworth filter), the proposed SwTDA finds spectral weights directly in the

frequency domain, which is more flexible in depicting the frequency characteristics.

## V. CONCLUSION

In this study, we introduce a method of Spectrum-weighted Tensor Discriminant Analysis (SwTDA) to motor imagery classification in BCI application. The SwTDA employs the technique of tensor-based discriminant analysis to solve for the combined optimization problem of spatial and spectral filters. The alternative and iterative learning method is adopted to optimize the cost function and derive spatial-spectral filters together with other patterns. Therefore, features extracted by the SwTDA can reflect the intrinsic characteristics in multiway data, which assists to enhance discriminability and robustness. Under the same condition, the experimental results of two EEG datasets demonstrate that the proposed method improves the classification accuracy in comparison with state-of-the-art methods. Furthermore, significant spatial-spectral patterns that are determined by the SwTDA are useful for studying the oscillatory patterns of different EEG tasks and contribute to a better understanding of the task mechanisms. Finally, since the extracted features are tensors, the combination of SwTDA and the tensor-based classifier (e.g., Riemannian geometry classifier and deep neural network) deserves to be explored for further improving the classification performance.

## APPENDIX A SOLUTION OF SwTDA

*Tensor Operation 1:* The mode- $n$  matricization of a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is denoted by  $X_{(n)} \in \mathbb{R}^{I_n \times (I_1 \times \dots \times I_{n-1} \times I_{n+1} \times \dots \times I_N)}$ . The element  $(i_n, j)$  of  $X_{(n)}$  is

$$X_{(n)}(i_n, j) = \mathcal{X}_{i_1, i_2, \dots, i_N} \quad (16)$$

where  $j = 1 + \sum_{k=1, k \neq n}^N (i_k - 1) J_k$ , with  $J_k = \prod_{m=1, m \neq n}^{k-1} I_m$ .

*Tensor Operation 2:* The contracted product of tensor  $\mathcal{X}$  along all but the mode- $n$  is defined as

$$\langle \mathcal{X}, \mathcal{X} \rangle_{-n} = \mathbf{X}_{(n)} \mathbf{X}_{(n)}^T \quad (17)$$

where  $\mathbf{X}_{(n)}$  denotes the mode- $n$  matricization of the tensor  $\mathcal{X}$ .

*Definition 1:* Concatenating a set of  $M$  tensor samples  $\{\mathcal{X}_m \in R^{I_1 \times I_2 \times \dots \times I_N}, m = 1, 2, \dots, M\}$  along the mode- $(N+1)$  in tensor  $\mathcal{X}_m \in R^{I_1 \times I_2 \times \dots \times I_N \times I_M}$  is denoted as

$$\mathcal{X} = \text{cat}(N+1, \mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_M) \quad (18)$$

The alternating method to solve for the problem (10) is detailed as follows. Supposing that  $\{\mathbf{U}, \mathbf{p}\}$  are given, we can represent  $\mathbf{S}_w$  and  $\mathbf{S}_b$  as

$$\begin{aligned} \mathbf{S}_w &= \sum_{m=1}^M \|[(\mathcal{X}_m - \bar{\mathcal{X}}_{k_m}) \times_1 \mathbf{U}^T \times_3 \mathbf{p}^T] \times_2 \mathbf{W}^T\|^2 \\ &= \text{tr}[\mathbf{W}^T \mathbf{S}_w^{(-2)} \mathbf{W}] \end{aligned} \quad (19)$$

$$\begin{aligned} \mathbf{S}_b &= M_k \sum_{m=1}^M \|[(\bar{\mathcal{X}}_k - \bar{\mathcal{X}}) \times_1 \mathbf{U}^T \times_3 \mathbf{p}^T] \times_2 \mathbf{W}^T\|^2 \\ &= \text{tr}[\mathbf{W}^T \mathbf{S}_b^{(-2)} \mathbf{W}] \end{aligned} \quad (20)$$

where  $\text{tr}[\cdot]$  returns the sum of the diagonal elements in the square matrix.  $\mathbf{S}_w^{(-2)}$  and  $\mathbf{S}_b^{(-2)}$  are the mode-2 scatter matrices, which can be expressed via tensor contracted products

$$\mathbf{S}_w^{(-2)} = \sum_{m=1}^M \langle \mathcal{A}_m, \mathcal{A}_m \rangle_{-2} = \langle \mathcal{A}, \mathcal{A} \rangle_{-2} \quad (21)$$

$$\mathbf{S}_b^{(-2)} = \sum_{k=1}^K \langle \mathcal{B}_m, \mathcal{B}_m \rangle_{-2} = \langle \mathcal{B}, \mathcal{B} \rangle_{-2} \quad (22)$$

where

$$\hat{\mathcal{X}}_m = \mathcal{X}_m - \bar{\mathcal{X}}_{k_m}, \quad \hat{\mathcal{X}} = \text{cat}(4, \hat{\mathcal{X}}_1, \hat{\mathcal{X}}_2, \dots, \hat{\mathcal{X}}_M) \quad (23)$$

$$\tilde{\mathcal{X}}_m = \sqrt{M_k}(\bar{\mathcal{X}}_k - \bar{\mathcal{X}}), \quad \tilde{\mathcal{X}} = \text{cat}(4, \tilde{\mathcal{X}}_1, \tilde{\mathcal{X}}_2, \dots, \tilde{\mathcal{X}}_M) \quad (24)$$

$$\mathcal{A}_m = \hat{\mathcal{X}}_m \times_1 \mathbf{U} \times_3 \mathbf{p} \quad (25)$$

$$\mathcal{B}_m = \tilde{\mathcal{X}}_m \times_1 \mathbf{U} \times_3 \mathbf{p} \quad (26)$$

$$\mathcal{A} = \hat{\mathcal{X}} \times_1 \mathbf{U} \times_3 \mathbf{p} \quad (27)$$

$$\mathcal{B} = \tilde{\mathcal{X}} \times_1 \mathbf{U} \times_3 \mathbf{p} \quad (28)$$

In above equations,  $M_k$  is the sample number of class  $k$ , and  $k_m$  is the class label for the  $m$ -th training sample.  $\bar{\mathcal{X}}_k$  and  $\bar{\mathcal{X}}_{k_m}$  denote the mean of the data in class  $k$  and class  $k_m$  respectively.  $\bar{\mathcal{X}}$  is the mean of all samples.  $\hat{\mathcal{X}}$ ,  $\hat{\mathcal{X}}_k$  and  $\hat{\mathcal{X}}_{k_m}$  are computed as

$$\bar{\mathcal{X}} = \frac{1}{M} \sum_{m=1}^M \mathcal{X}_m \quad (29)$$

$$\hat{\mathcal{X}}_k = \frac{1}{M_k} \sum_{m \in k} \mathcal{X}_m, \quad \hat{\mathcal{X}}_{k_m} = \frac{1}{M_{k_m}} \sum_{m \in k_m} \mathcal{X}_m \quad (30)$$

So when fixing  $\mathbf{U}$  and  $\mathbf{p}$ , the learning rule for  $\mathbf{W}$  can be found via solving the below optimization

$$\begin{aligned} \mathbf{W} &= \arg \max \frac{\text{tr}[\mathbf{W}^T \mathbf{S}_b^{(-2)} \mathbf{W}]}{\text{tr}[\mathbf{W}^T \mathbf{S}_w^{(-2)} \mathbf{W}]}, \\ \text{subject to } \mathbf{W}^T \mathbf{W} &= \mathbf{I} \end{aligned} \quad (31)$$

where  $\mathbf{W}^T \mathbf{W} = \mathbf{I}$  is to ensure the orthonormality of the projection matrix. The problem of equation (31) can be solved by generalized eigenvalue problem

$$\mathbf{S}_b^{(-2)} \mathbf{W} = \mathbf{S}_w^{(-2)} \mathbf{W} \Lambda \quad (32)$$

where  $\mathbf{W}$  is the eigenvectors corresponding to the largest  $J$  eigenvalues in the diagonal matrix  $\Lambda$ . Similar way can be used to solve for  $\mathbf{U}$  or  $\mathbf{p}$ .

## ACKNOWLEDGMENT

The authors would like to thank the Berlin BCI Group and W. B. Yi for providing the EEG data set used in this article. They also like to thank Soroosh Shahtalebi from the Concordia University for helpful discussion of the ECCSP-TB2B method.

## REFERENCES

- [1] R. Chai, S. H. Ling, G. P. Hunter, Y. Tran, and H. T. Nguyen, "Brain-computer interface classifier for wheelchair commands using neural network with fuzzy particle swarm optimization," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1614–1624, Sep. 2014.
- [2] P. Arico, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, and F. BabILONI, "Passive BCI in operational environments: Insights, recent advances, and future trends," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1431–1436, Jul. 2017.
- [3] L. Yao, X. Sheng, D. Zhang, N. Jiang, N. Mrachacz-Kersting, X. Zhu, and D. Farina, "A stimulus-independent hybrid BCI based on motor imagery and somatosensory attentional orientation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 9, pp. 1674–1682, Sep. 2017.
- [4] M. Jalipour Monesi and S. Hajipour Sardouie, "Extended common spatial and temporal pattern (ECSTP): A semi-blind approach to extract features in ERP detection," *Pattern Recognit.*, vol. 95, pp. 128–135, Nov. 2019.
- [5] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "EEG feature comparison and classification of simple and compound limb motor imagery," *J. Neuroeng. Rehabil.*, vol. 10, no. 1, p. 106, 2013.
- [6] F. Yger, M. Berar, and F. Lotte, "Riemannian approaches in brain-computer interfaces: A review," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1753–1762, Oct. 2017.
- [7] F. Lotte, "A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update," *J. Neural Eng.*, vol. 15, no. 3, 2018, Art. no. 031005.
- [8] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Dec. 2000.
- [9] W. Wu, S. Jiang, R. Liu, W. Jin, and C. Ma, "Economic development, demographic characteristics, road network and traffic accidents in zhongshan, China: Gradient boosting decision tree model," *Transportmetrica A, Transp. Sci.*, vol. 16, no. 3, pp. 359–387, Jan. 2020.
- [10] M. Hamed, S.-H. Salleh, and A. M. Noor, "Electroencephalographic motor imagery brain connectivity analysis for BCI: A review," *Neural Comput.*, vol. 28, no. 6, pp. 999–1041, Jun. 2016.
- [11] V. Jayaram, M. Alamgir, Y. Altun, B. Scholkopf, and M. Grosse-Wentrup, "Transfer learning in brain-computer interfaces," *IEEE Comput. Intell. Mag.*, vol. 11, no. 1, pp. 20–31, Feb. 2016.
- [12] J. H. Li, C. Li, and A. Cichocki, "Canonical polyadic decomposition with auxiliary information for brain-computer interface," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 263–271, Jan. 2017.

- [13] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, “The non-invasive Berlin brain–computer interface: Fast acquisition of effective performance in untrained subjects,” *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.
- [14] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b,” *Frontiers Neurosci.*, vol. 6, p. 39, Mar. 2012.
- [15] Y. Zhang, G. X. Zhou, J. Jin, X. Y. Wang, and A. Cichocki, “Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface,” *J. Neurosci. Methods*, vol. 255, pp. 85–91, Nov. 2015.
- [16] A. S. Aghaei, M. S. Mahanta, and K. N. Plataniotis, “Separable common spatio-spectral patterns for motor imagery BCI systems,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 1, pp. 15–29, Jan. 2016.
- [17] H.-I. Suk and S.-W. Lee, “A novel Bayesian framework for discriminative feature extraction in brain–computer interfaces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.
- [18] W. Yi, S. Qiu, K. Wang, H. Qi, L. Zhang, P. Zhou, F. He, and D. Ming, “Evaluation of EEG oscillatory patterns and cognitive process during simple and compound limb motor imagery,” *PLoS ONE*, vol. 9, no. 12, 2014, Art. no. e114853.
- [19] S. Shahalebi and A. Mohammadi, “Bayesian optimized spectral filters coupled with ternary ECOC for single-trial EEG classification,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 12, pp. 2249–2259, Dec. 2018.
- [20] D. Tao, X. Li, X. Wu, and S. J. Maybank, “General tensor discriminant analysis and Gabor features for gait recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
- [21] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, “Spatial-temporal discriminant analysis for ERP-based brain–computer interface,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 2, pp. 233–243, Mar. 2013.
- [22] A. H. Phan and A. Cichocki, “Tensor decompositions for feature extraction and classification of high dimensional datasets,” *Nonlinear Theory Appl., IEICE*, vol. 1, no. 1, pp. 37–68, 2010.
- [23] F. Qi, Y. Li, and W. Wu, “RSTFC: A novel algorithm for spatio-temporal filtering and classification of single-trial EEG,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3070–3082, Dec. 2015.
- [24] W. Wu, X. Gao, B. Hong, and S. Gao, “Classifying single-trial EEG during motor imagery by iterative spatio-spectral patterns learning (ISSPL),” *IEEE Trans. Biomed. Eng.*, vol. 55, no. 6, pp. 1733–1743, Jun. 2008.
- [25] H. Higashi and T. Tanaka, “Simultaneous design of FIR filter banks and spatial patterns for EEG signal classification,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 4, pp. 1100–1110, Apr. 2013.
- [26] J. Meng, L. Yao, X. Sheng, D. Zhang, and X. Zhu, “Simultaneously optimizing spatial spectral features based on mutual information for EEG classification,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 1, pp. 227–240, Jan. 2015.
- [27] A. K. Jain, R. P. W. Duin, and J. Mao, “Statistical pattern recognition: A review,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2002.
- [28] S. Selim, M. M. Tantawi, H. A. Shedeed, and A. Badr, “A CSPAM-BA-SVM approach for motor imagery BCI system,” *IEEE Access*, vol. 6, pp. 49192–49208, 2018.
- [29] F. Lotte and C. Guan, “Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms,” *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.
- [30] W. Samek, M. Kawanabe, and K.-R. Müller, “Divergence-based framework for common spatial patterns algorithms,” *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.
- [31] D. Thiyam, S. Cruces, J. Olias, and A. Cichocki, “Optimization of alpha-beta log-det divergences and their application in the spatial filtering of two class motor imagery movements,” *Entropy*, vol. 19, no. 3, p. 89, 2017.
- [32] H. Cecotti, M. P. Eckstein, and B. Giesbrecht, “Single-trial classification of event-related potentials in rapid serial visual presentation tasks using supervised spatial filtering,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 11, pp. 2030–2042, Nov. 2014.
- [33] Y. R. Tabar and U. Halici, “A novel deep learning approach for classification of EEG motor imagery signals,” *J. Neural Eng.*, vol. 14, no. 1, Feb. 2017, Art. no. 016003.
- [34] S. Sakhavi, C. Guan, and S. Yan, “Learning temporal information for brain–computer interface using convolutional neural networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5619–5629, Nov. 2018.
- [35] J. Hu and J. Min, “Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model,” *Cognit. Neurodyn.*, vol. 12, no. 4, pp. 431–440, Aug. 2018.
- [36] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, “Transfer learning: A Riemannian geometry framework with applications to brain–computer interfaces,” *IEEE Trans. Biomed. Eng.*, vol. 65, no. 5, pp. 1107–1116, May 2018.
- [37] H. Woehrle, M. M. Krell, S. Straube, S. K. Kim, E. A. Kirchner, and F. Kirchner, “An adaptive spatial filter for user-independent single trial detection of event-related potentials,” *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1696–1705, Jul. 2015.



**SHOULIN HUANG** received the M.S. degree in signal and information processing from the Guilin University of Electronic Technology, Guilin, China, in 2009. He is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China. His current interests include biomedical signal processing and brain–computer interface.



**YANG CHEN** received the M.S. degree in communication engineering from the Harbin Institute of Technology at Shenzhen, Shenzhen, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Electronics and Information Engineering. His research interests include health informatics, and biomedical signal processing and modeling.



**TONG WANG** (Member, IEEE) received the B.Eng. degree in electrical engineering and automation from the Beijing University of Aeronautics and Astronautics (now Beihang University), Beijing, China, in 2006, and the M.Sc. degree (Hons.) in communications engineering and the Ph.D. degree in electronic engineering from the University of York, York, U.K., in 2008 and 2012, respectively. From 2012 to 2015, he was a Research Associate with the Institute for Theoretical Information Technology, RWTH Aachen University, Aachen, Germany. From 2014 to 2015, he was a Research Fellow of the Alexander von Humboldt Foundation. Since March 2016, he has been with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China, where he is currently an Assistant Professor. His research interests include sensor networks, cooperative communications, adaptive filtering, and resource optimization.



**TING MA** (Member, IEEE) received the B.Eng. and M.Sc. degrees from the Harbin Institute of Technology ( HIT ), Harbin, China, in 1999 and 2001, respectively, and the Ph.D. degree in biomedical engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in 2004. From 2004 to 2009, she was a Postdoctoral Fellow with CUHK. From 2014 to 2016, she was a Visiting Professor with Johns Hopkins University. Since June 2009, she has been with the School of Electronics and Information Engineering, Harbin Institute of Technology at Shenzhen, Shenzhen, China, where she is currently an Associate Professor. Her research interests include neural computing, neural image processing, and brain informatics.